D.1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 November 2002 (28.11.2002)

PCT

(10) International Publication Number
WO 02/096014 A1

(51) International Patent Classification⁷: H04L 9/00

(21) International Application Number: PCT/US02/15948

(22) International Filing Date: 21 May 2002 (21.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/291,944 21 May 2001 (21.05.2001) US

(71) Applicant *(for all designated States except US)*: **FORMATTA** [US/US]; 7918 Jones Branch Avenue, McLean, VA 22101 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **WERNET, Paul, G.** [US/US]; 4560 King Edward Court, Annandale, VA 22003 (US). **WHITMORE, Dean, Joseph** [US/US]; 1201 Braddock Place #813, Alexandria, VA 22314 (US).

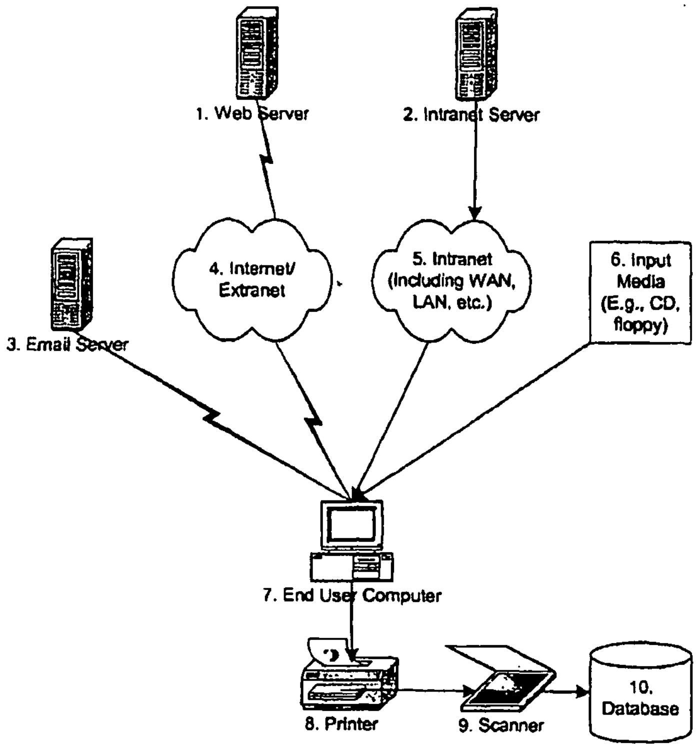
(74) **Agent: MUTTER, Michael, K.**; Birch, Stewart, Kolasch & Birch, LLP, P.O. Box 747, Falls Church, VA 22040-0747 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR INCREASING THE ACCURACY AND SECURITY OF DATA CAPTURE FROM A PAPER FORM



(57) Abstract: A system and method dynamically generates and prints an encrypted two-dimensional barcode on a electronic form intended for data entry. A field set is specified to identify the data that will be encrypted within the two-dimensional barcode. A password and specified and is used to encrypt the data; and form identification information is also provided identifying the form electronically when it is submitted and scanned (9). A two-dimensional barcode is dynamically generated when an end user (7) prints the form(8), such that the data content of the two-dimensional barcode is the data entered into the form by the end user. The two-dimensional barcode is printed with the form, such that when the user requests printing, the data and field identification information from each field in the selected field set is extracted, the data and field identification information is encrypted with the specified encryption password, the identifying information is included, but not encrypted, and the printed two-dimensional barcode contains both the encrypted information and the unencrypted identifying information.

European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD AND SYSTEM FOR INCREASING THE ACCURACY AND SECURITY OF DATA CAPTURE FROM A PAPER FORM

FIELD OF THE INVENTION

This invention relates in general to computer software, and in particular to a method and system for increasing the accuracy and security of data capture from a paper form where the form was completed electronically and printed out prior to submission. The invention utilizes two-dimensional barcode technology to dynamically capture data entered electronically into the form. The data contained in this two-dimensional barcode is then encrypted for security and authentication purposes and printed on the form when the user prints the form. When the printed form is received by the Data Collector at a central processing site, the two-dimensional barcode is scanned and decrypted, the form is authenticated, and the data is extracted virtually error-free.

BACKGROUND OF THE INVENTION

Electronic forms applications consist of three primary components: design software for the Form Author, filler software for the End-User completing the form, and server software for the Form Distributor and/or Data Collector (the Form Distributor and the Data Collector may or may not be the same entity, and either may or may not be related to the Form Author).

The design software is used to create the electronic form (e-form), or user interface of the data container, as well as the algorithms associated with the e-form and data to be entered into the e-form. The Form Author may design the e-form as a traditional electronic form or integrate elements of hypertext markup language (HTML), extensible markup language (XML), portable document format (PDF), graphic elements (e.g., GIF, TIF. JPEG) and other objects to achieve the desired user interface. The designer may also specify data edits, validation, and other functions such as encryption, glyph generation, e-mail routing information, etc. that govern the behavior of the e-form in the filler application.

Filler software allows End Users to view and interact with the e-forms created using the design software. User interactions include filling out the e-form electronically, saving the e-form, printing the e-form, submitting the e-form, and similar functions depending on the algorithms associated with the e-form by the designer.

Server software allows form distributors and Data Collectors to process forms (e-forms and paper forms) automatically. For e-forms, the server software enables the Form Distributor to pre-fill forms with data from a database and distribute the pre-filled forms to End Users electronically (e.g., via email). Optionally, the distributor may encrypt the pre-filled data, or subsets of the pre-filled data, prior to distributing the e-forms. Server software also enables Data Collectors to process incoming e-forms electronically and automatically. An example of such processing would be to receive the incoming e-form, identify the form, authenticate the form, decrypt the form, extract the data from the form, and write the data to a database. For paper forms, the server software enables Data Collectors to automatically extract the form data from the paper form by scanning a two-dimensional barcode containing the form data, decrypt the data extracted from the bar code, authenticate the form, and write the extracted data to a database.

Prior to this invention, if a Data Collector required the End User to submit the form on paper (as is the case if the form requires a 'wet' signature), the Data Collector had to rely on OCR/ICR/OMR, re-keying, or some other method to extract the data from the paper form. These data extraction methods are prone to transcription errors, are costly, and cannot detect counterfeit forms. This invention allows Data Collectors to receive the printed form (with the 'wet' signature) and extract the data by scanning a two-dimensional barcode printed on the form. This method is more accurate than prior data extraction methods, because scanning a two-dimensional barcode is an all-or-nothing proposition: either it scans correctly, and the data is extracted exactly as it was entered into the form; or it doesn't scan at all, so no data errors are introduced via the scanning process (the form would have to go to exception processing instead). It is also more secure, since the data in the printed two-dimensional barcode is encrypted, ensuring that only an authorized party (such as the Data Collector) can extract the data electronically.

Counterfeit forms are not a new concept, but their likely frequency and the damage they can wreak on Data Collectors are dramatically increased in the world of PC-rendered paper forms (i.e., where forms are obtained electronically by an End User and printed out by the End User before submission to the Data Collector). This scenario presents risks to the Data Collector, since a knowledgeable End User could conceivably alter a form before submitting it (either electronically or printed on paper). For example, using form design tools, an End User could change the perjury statement common to many forms to read as follows: “I do NOT declare under penalty of perjury…”. The simple insertion of the word NOT in the perjury statement clearly violates the intention of the Data Collector. It then becomes a further obligation on the Data Collector to validate the authenticity of the submitted forms themselves, not just the data included on those forms. However, with this invention, the data imbedded in the two-dimensional bar code can only be successfully decrypted by an entity with the correct Form Lock password, or keyset. If the data cannot be decrypted with the correct Form Lock password, then the Data Collector or other authorized entity knows the form itself is counterfeit. The same is true for an electronic form submission, since the Data Collector will only be able to decrypt the data on the form if the original Form Lock password functions.

SUMMARY OF THE INVENTION

The present invention disclosed herein comprises a method and system for increasing the accuracy and security of data capture from a paper form where the form was completed electronically and printed out on paper prior to submission, and in authenticating the printed form. The invention utilizes two-dimensional barcode technology to dynamically capture data entered electronically into the form. This two-dimensional barcode is then encrypted for security and printed on the form when the user prints the form. When the printed form is received by the Data Collector, the two-dimensional barcode is scanned and decrypted, and the data is extracted virtually error-free, eliminating the need for more costly, less efficient data extraction technologies and techniques. Successful decryption of the data authenticates the form as well, since the decryption will fail if the form has been altered or otherwise tampered with.

DETAILED DESCRIPTION OF THE DRAWINGS

Referring to FIG. **1**, there is depicted a graphic representation of a data processing system which may be utilized to implement the present invention. As may be seen, data processing system may include a plurality of networks, such as Local Area Network (LAN), Wide Area Network (WAN) and Internet, each of which may include a plurality of individual computers respectively. Those skilled in the art will appreciate that a plurality of workstations coupled to a host processor may be utilized for each such network. As is common in such data processing systems, each individual computer may be coupled to a storage device and/or printer/output device and/or input device.

The data processing system may also include multiple server computers, such as mainframe computer, which may be coupled to computer, LAN, WAN or Internet by means of communications link. The server computers may also be coupled to a storage device which may serve as remote storage for the End User computer, LAN, WAN or Internet. Similarly, the End User computer, WAN and Internet may be coupled via communications link through a subsystem control unit/communications controller and communications link to a gateway server creating an inter-network link.

With respect to the End User computer, LAN, WAN and Internet, a plurality of documents or resource objects may be stored within storage device and controlled by a server computer, as resource manager or library service for the resource objects thus stored. Those skilled in the art will appreciate that the server computer may be located a great geographic distance from LAN and similarly, LAN may be located a substantial distance from the End User computer. For example, the End User computer may be located in Colorado while LAN may be located in Washington and server computer may be located in New York.

Software program code which employs the present invention is typically stored in the memory of a storage device of a stand alone work station or storage device of a server computer from which a developer may access the code. For distribution purposes, the software program code may be embodied on any of a variety of known media for use with a data processing system, such as a diskette or CD-ROM or may be distributed to users from the memory of one computer system over a network of some type to other computer systems for use by users of such other systems. Such techniques and methods for embodying software code on media and/or distributing software code are well known and will not be further discussed herein.

With respect to the present invention, the End User uses the End User computer **7** to access the e-form. The End User computer is running software program code which employs the present invention. The e-form is accessed directly from a storage device connected to the End User computer 7, such as a local hard drive, or from some form of input media **6**, or from an email message from an email server **3**, or from a communications link to the Internet/Extranet **4** and web server **1**, or from a local Intranet **5** and Intranet server **2** or some similar access method. The End User uses the End User computer **7** to view the eform, fill in data in the data fields, save the form, and to perform other similar actions. When the eform is completed and the End User prints the eform to the printer **8**, the End User computer **7** carries out the instructions in the software program code which employs the present invention, dynamically creates the encrypted data set from the field data contained in the eform and prints the encrypted data along with identifying information in a two-dimensional barcode on the paper document. The printed document is then sent to the Data Collector. When the Data Collector receives the printed document, the document is scanned using scanner **9** which is attached to a scanning station. The scanning station is running software program code which employs the present invention to decipher the two dimensional barcode and process the data appropriately (identifying the originating form, authenticating the form, decrypting the encrypted data set, identifying the correct database, and writing the data to the database **10**). Once processed, the data may be written to a database **10** or some other storage device or passed to another system or application for continued processing or other purposes.

DETAILED DESCRIPTION OF THE INVENTION

When the e-form is being designed using the designer software, the designer selects a set of fields on the e-form for creating the dynamic two-dimensional barcode (the 2D barcode field set). This set of fields can include all of the fields on the form or only a selected subset of fields on the form. The designer also selects an encryption password that will be used to encrypt the data in the 2D barcode field set before it is printed on the paper form as a two-dimensional barcode. An example of a common two-dimensional barcode, the PDF417, appears below:

When the End User opens the e-form in the filler software, the End User can enter data via PC keyboard and mouse selections into the data fields electronically (e.g., type in the data; click on pull-down menus to select specific item(s)). When the user is done filling out the form, the user can print the form. When the user clicks the print button (or icon), the filler software automatically extracts the data from the 2D barcode field set, encrypts the data, and prints a two-dimensional barcode containing the encrypted data onto the paper form. The printed two-dimensional barcode also contains some unencrypted data (e.g., form ID number, registration number, or similar identifying information) that is used to identify the form when it is returned to the Data Collector. This process is transparent to the End User, except that the End User will see a two-dimensional barcode printed on the form. The End User may then submit the printed form to the Data Collector.

When the Data Collector receives the printed form, the Data Collector may scan the two-dimensional barcode printed on the form to extract the form data electronically. This is done using a conventional two-dimensional barcode scanner and the server software. The scanner scans the barcode and extracts the encrypted field data from the barcode. The server software identifies the originating e-form based on non-encrypted data included in the two-dimensional barcode. The server software then applies the registered keyset associated with the e-form identification information to decrypt the encrypted field data. Successful decryption of the dataset in the barcode authenticates the form. After the data is decrypted it is written to a database.

We claim:

1. A method of dynamically generating and printing an encrypted two-dimensional barcode containing the data entered into an electronic form, which comprises:

- specifying the field set to identify the data that will be encrypted and contained in the two-dimensional barcode;
- specifying the names of the fields in the field set, so that the fields can be made to correspond to fields in a database associated with the form;
- including the field names or other identifying information along with the field content for each selected field in the field set, such that each field name and the respective field content are and remain associated with each other;
- specifying the encryption password that will be used to encrypt the data;
- specifying the form identification information, that will be included in the two-dimensional barcode but not encrypted, to identify the form electronically when it is submitted and scanned;
- dynamically generating the two-dimensional barcode when the End User prints the form, such that the data content of the two-dimensional barcode is the data entered into the form by the End User;
- dynamically generating the two-dimensional barcode when the End User prints the form, such that when the user clicks the print button, the data and field identification information from each field in the selected field set is extracted, the data and field identification information is encrypted with the specified encryption password, the identifying information is included, but not encrypted, and the two-dimensional barcode containing both the encrypted information and the unencrypted identifying information is printed on the form.

2. A method of automatically extracting data from a printed, encrypted two-dimensional barcode, which comprises the steps of:

- scanning the two-dimensional barcode with a conventional two-dimensional barcode scanner;
- identifying the originating electronic form based on the unencrypted identifying information included in the two-dimensional barcode;

- retrieving the specified encryption keyset to decrypt the data contained in the two-dimensional barcode based on properly identifying the originating electronic form;
- decrypting the encrypted data contained in the two-dimensional barcode using the specified encryption password;
- authenticating the dataset as having originated from the originating form based on the success or failure of the decryption process;
- identifying the correct database to write the data to based on properly identifying the originating electronic form, which is associated with a specific database;
- writing the decrypted data to the specified database, such that the field data from the form is written to the appropriate field(s) in the database as defined by the field names from the form and the associated database fields.

3. A method of authenticating a printed version of an electronic form from a printed, encrypted two-dimensional barcode, which comprises the steps of:

- scanning the two-dimensional barcode with a conventional two-dimensional barcode scanner;
- identifying the originating electronic form based on the unencrypted identifying information included in the two-dimensional barcode;
- retrieving the specified encryption keyset to decrypt the data contained in the two-dimensional barcode based on properly identifying the originating electronic form;
- decrypting the encrypted data contained in the two-dimensional barcode using the specified encryption password;
- authenticating the printed form as the printed product of the original electronic form based on the success or failure of the decryption process.

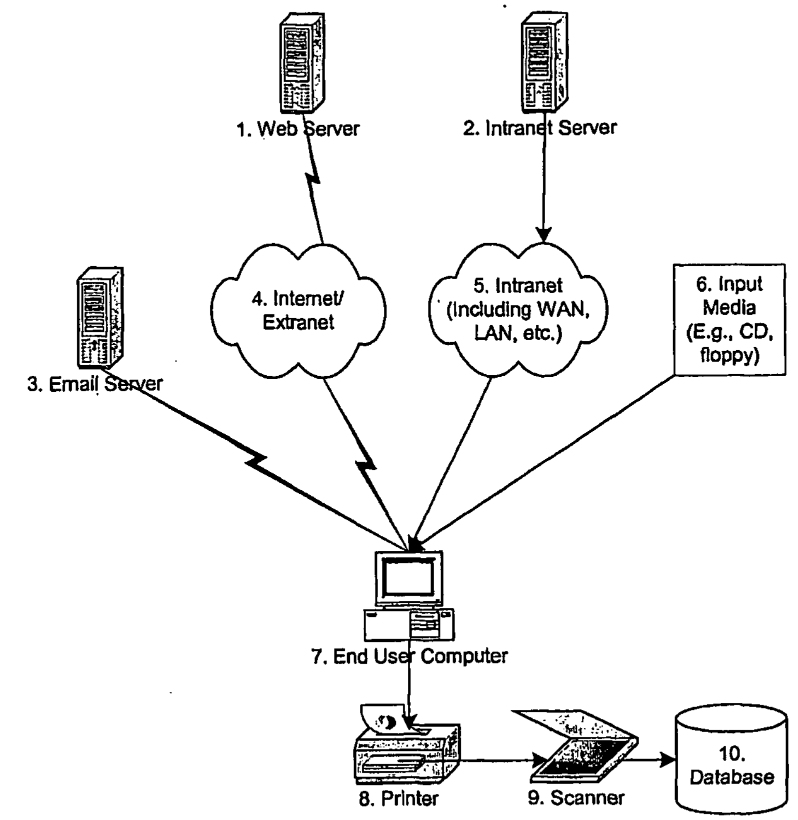

1. Web Server
2. Intranet Server
3. Email Server
4. Internet/ Extranet
5. Intranet (Including WAN, LAN, etc.)
6. Input Media (E.g., CD, floppy)
7. End User Computer
8. Printer
9. Scanner
10. Database

Figure 1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/15948

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :H04L 9/00
US CL : 380/51; 705/51,60,401
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 380/51; 705/51,60,401

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6,005,945 A(WHITEHOUSE) 21 December 1999, col.2,lines 43-47, col.4,lines 5-32,col.6,lines 31-65.	1-3
Y	US 5,598,477 A(BERSON) 28 January 1997, col.4,lines 61-67, col.5,lines 13-45.	1-3
A	US 5,864,622 A(MARCUS) 26 January 1999, col.4,lines 6-43, col.5,lines 8-37.	1-3
A	US 6,085,976 A(SEHR) 11 June 2000, col.17, lines 5-42, col.31,lines 5-32	1-3

□ Further documents are listed in the continuation of Box C. □ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 28 JUNE 2002	Date of mailing of the international search report 02 AUG 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer GAIL HAYES Telephone No. (703) 305-0042

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/15948

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

STN, EAST
search terms: postage,barcode,encrypt,encipher,cipher,ticket,electronic form,password,key,print,generate